

w 1147

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-196989
(43)Date of publication of application : 12.07.2002

(51)Int.CI. G06F 12/16
G06F 3/06

(21)Application number : 2001-075365
(22)Date of filing : 16.03.2001

(71)Applicant : HITACHI LTD
(72)Inventor : SUZUKI TORU
NAKAMURA KATSUNORI
NAGASAWA MITSUO
KIMURA YUKIHISA
KOIDE TAKESHI

(30)Priority

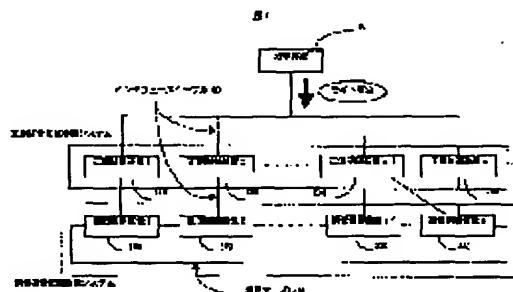
Priority number : 2000316743 Priority date : 17.10.2000 Priority country : JP

(54) DATA DUPLEX METHOD AND DEVICE FOR STORAGE SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To asynchronously and consistently copy data from a plurality of original-side control devices to a plurality of duplicate-side control devices.

SOLUTION: All of the duplicate-side control devices are connected together into a loop, the respective duplicate-side control devices determine guarantee time by circulating on the looped communication passage in a time applied to data received from the original-side control devices, and circulates on the looped communication passage in the determined guarantee time.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

W 1149

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号
特開2002-196989
(P2002-196989A)

(43)公開日 平成14年7月12日(2002.7.12)

審査請求 未請求 請求項の数14 OL (全 18 頁)

(21)出願番号	特願2001-75365(P2001-75365)
(22)出願日	平成13年3月16日(2001.3.16)
(31)優先権主張番号	特願2000-316743(P2000-316743)
(32)優先日	平成12年10月17日(2000.10.17)
(33)優先権主張国	日本(JP)

(71) 出願人 000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目 6 番地

(72) 発明者 鈴木 亨
神奈川県小田原市国府津2880番地 株式会社日立製作所ストレージシステム事業部内

(72) 発明者 中村 勝憲
神奈川県小田原市国府津2880番地 株式会社日立製作所ストレージシステム事業部内

(74) 代理人 100075096
弁理士 作田 康夫

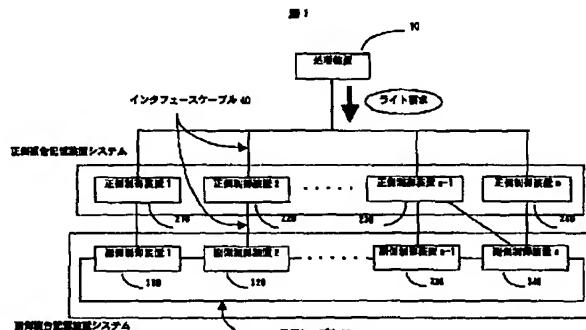
最終頁に続く

(54) 【発明の名称】 記憶装置システムのデータ二重化方法および装置

(57) 【要約】

【課題】 複数の正側制御装置から、複数の副側制御装置へ、一貫性のとれたデータの非同期コピーを可能とする。

【解決手段】 全副側制御装置をループ接続し、各副側制御装置は、正側制御装置から受信したデータに付与された時刻を前記ループ通信路で巡回することで保証時刻を決定し、決定した保証時刻をループ通信路で巡回する。



【特許請求の範囲】

【請求項1】正システムを構成する1台以上の正側制御装置から副システムを構成する1台以上の副側制御装置へデータのバックアップを行う記憶装置システムのデータ二重化方法であって、処理装置から前記正側制御装置にライト要求があった場合に、前記ライト要求のデータにライト時刻に関する時刻情報を付与して前記正側制御装置内のキャッシュメモリに格納するステップと、前記格納された前記ライトデータと前記時刻情報を前記副側制御装置に送るステップと、前記副側制御装置内のキャッシュメモリに前記ライトデータと前記時刻情報を格納するステップと、前記副側制御装置内の前記キャッシュメモリに格納された前記時刻情報を、前記副側制御装置間を互いに接続した通信路を通して他の副側制御装置に送るステップと、第1の正側制御装置から受け取った第1の副側制御装置の時刻情報と、第2の副側制御装置から送られた時刻情報を比較して、古い方の時刻情報を前記通信路を通して次の副側制御装置に送るステップと、及び、前記古い方の時刻情報を前記複数の副側制御装置間を一巡させて、前記複数の副側制御装置間に前記ライトデータが二重化されたことを保証するデータ保証時刻を決定するステップと、を備える記憶装置システムにおけるデータの二重化方法。

【請求項2】前記副側制御装置が、前記通信路で接続された他の副側制御装置に前記データ保証時刻を通知するステップと、及び、

前記複数の副側制御装置間で前記データ保証時刻を前記通信路で一巡させるステップとを備える請求項1記載の記憶装置システムのデータ二重化方法。

【請求項3】前記副側制御装置が、前記通信路で接続された他の副側制御装置に前記古い方の時刻情報と前記データ保証時刻とを一緒にして送信し、一巡させ、前記複数の副側制御装置間にデータ保証時刻の決定と複数の副側制御装置間での前記データ保証時刻の通知を合せて行う請求項2記載の記憶装置システムのデータ二重化方法。

【請求項4】前記副側制御装置が、前記データ保証時刻の決定と前記データ保証時刻の通知の一巡とを、任意の時間間隔で複数個並行して実行するステップを備える請求項3記載の記憶装置システムのデータ二重化方法。

【請求項5】処理装置が書き込むデータを受信する複数の正側制御装置から、複数の副側制御装置へ前記データを二重化する方法であって、

前記複数の正側制御装置が、前記処理装置からデータを受信する第1ステップと、前記複数の正側制御装置が、前記データを処理装置から受領した時刻に関する第1の情報と前記データとを前記複数の副側制御装置へ送信する第2ステップと、及び、前記複数の副側制御装置が、

前記第1の情報に関する第2の情報を、前記複数の副側制御装置間で巡回通信する第3ステップとを有するデータ二重化方法。

【請求項6】前記第3ステップは、前記複数の副側制御装置をリング状に接続する物理的通信路上で実行される請求項5記載のデータ二重化方法。

【請求項7】前記第3ステップにおいて、第1の副側制御装置が送信した前記第2の情報は、前記第1の情報を受信した第2の副側制御装置によって、前記第1の情報を参照して更新される請求項5記載のデータ二重化方法。

【請求項8】前記第2の情報を送信した前記第1の副側制御装置は、前記巡回通信により第2の情報を受信したとき、受信した前記第2の情報から、前記複数の副側制御装置が前記データの二重化を保証する時間を決定する請求項7記載のデータ二重化方法。

【請求項9】前記第1の副側制御装置は、前記保証した時間を、前記巡回通信により、前記複数の副側制御装置へ通知する第4ステップとを有する請求項8記載のデータ二重化方法。

【請求項10】前記第3ステップと前記第4ステップを合せて実行する請求項9記載のデータ二重化方法。

【請求項11】前記第3ステップと前記第4ステップを合せたステップが、前記複数の副側制御装置をリング状に接続する物理的通信路上で、複数実行される請求項9記載のデータ二重化方法。

【請求項12】正システムを構成する1台以上の正側制御装置から副システムを構成する1台以上の副側制御装置へデータのバックアップを行う記憶装置システムにおけるデータ二重化方法であって、

前記副側制御装置間を通信手段で互いに接続するステップと、及び、前記副側制御装置間で前記通信手段を通してデータの管理情報を巡回して通信し、データに関する共通の管理情報を決定するステップとを有する記憶装置システムにおけるデータ二重化方法。

【請求項13】処理装置が書き込むデータを受信する複数の正側制御装置と、前記データを二重化データとして受信する複数の副側制御装置を有する記憶システムであって、

前記処理装置からデータを受信する前記複数の正側制御装置内の複数のデータ受信部と、

前記データを処理装置から受領した時刻に関する第1の情報と前記データとを前記複数の副側制御装置へ送信する、前記複数の正側制御装置内の複数のデータ送信部と、

前記第1の情報に関する第2の情報を、第1の副側制御装置から第2の副側制御装置へ送信する前記第1の副側制御装置内のデータ送信部と、及び、

前記送信された前記第2の情報を受信する前記第2の副

側制御装置内のデータ受信部とを有し、ここで、前記複数の副側制御装置は巡回通信路を形成する記憶システム。

【請求項14】前記巡回通信路は、物理的にリング形状の通信路である請求項13記載の記憶システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、処理装置（ホストコンピュータ）に接続された正側制御装置と正側制御装置に接続された副側制御装置との間でデータの2重化（データのバックアップ）を行う記憶装置システムに関する。

【0002】

【従来の技術】正・副の制御装置間で、データの2重化を行う方法として、正・副の制御装置をインターフェースケーブルで接続し、正側の制御装置が処理装置から受領したライト要求を副側の制御装置に発行することでデータの2重化を行う方法が一般的に知られている。

【0003】更に、正・副のシステムが各々複数（M, N台）の制御装置を備え、処理装置からM台の正側制御装置にライト要求が発行されるシステムで、データの2重化を行う方法も知られている。この方法については、例えば、特開平11-85408公報に開示がある。

【0004】この正・副のシステムが各々複数（M, N台）の制御装置を備える場合、各正側制御装置は処理装置（ホスト）から受領したライト要求に、ライト要求が発行されたライト時刻を付与して、インターフェースケーブルで接続された各副側制御装置にライト要求を発行し、副側制御装置にライトデータとライト時刻を送信する。一方、副側システムでは、複数の副側制御装置の中から1台をマスタ制御装置と設定され、その設定されたマスタ制御装置とその他の副側制御装置が、各々データ転送路によって1対1接続されている。そして、マスタ以外の副側制御装置は、正側制御装置から受領したライトデータのライト時刻（ホストから正側制御装置へのライト時刻）に関する情報を、副側のマスタ制御装置にデータ転送路で送信する。マスタ制御装置は、全ての副側制御装置から受領したライト時刻に関する情報（マスタ制御装置自身のライト時刻情報も含む）を参照して保証時刻を決定する。ここで、保証時刻とは、「全副側制御装置において、保証時刻よりも過去のライト時刻（ホストから正側へのライト時刻）が付与されているライトデータについては全て副側制御装置で受領済みであると保証できる時刻」である。即ち、決定された保証時刻は、保証時刻以前に処理装置から各正側制御装置に発行された全ライト要求に関する全ライトデータが副側制御装置で受領済み（2重化済み）であることを保証する。マスタ制御装置は、この決定した保証時刻を全副側制御装置に通知し、各副側制御装置は、保証時刻以前のライト時刻を持つライトデータの2重化が保証されたことを確認

できる。そして、この保証時刻を知らされた制御装置は、例えば、正側のデータが何らかの障害で破壊された場合に、この保証時刻以前に正側制御装置から受領したデータを利用することができるうことになる。

【0005】

【発明が解決しようとする課題】上記の方法では、正・副のシステムが複数の制御装置を備えるシステムにおいて、データの2重化を保証し、この副側のデータを使用するためには、保証時刻を決定する必要がある。そして、保証時刻を決定するためには、副側のマスタ制御装置とその他の各副側制御装置との間で、各副側制御装置がライト時刻に関する情報をマスタ制御装置に送信するのと、前記マスタ制御装置が決定した保証時刻を各副側制御装置に通知するので、マスタ副制御装置以外の副制御装置は各々2回の通信が必要であった。

【0006】また上記の方法では、データの2重化を保証するための処理・通信がマスタ制御装置に集中してしまうため、特に制御装置を多数備えるようなシステムにおいては、マスタ制御装置における、通信性能と処理性が低下してしまう課題があった。

【0007】更に上記方法では、マスタ制御装置に、他の（N-1）台の副側制御装置と個々に通信が行える機能及び物理的な構成を持たせなければならず、特に副システムを構成する制御装置の数が多数に及ぶ場合には、マスタ制御装置にその他の全ての副側制御装置と通信を行う機能及び物理的な構成を持たせることは困難となっていた。

【0008】更に上記方法では、マスタ制御装置に障害が発生した場合を考慮して、マスタ制御装置としての機能を副側制御装置に分散させ、冗長性を持たせることを考えると、マスタ制御装置としての機能及び物理的な構成を各副側制御装置に分散させて持たせる必要があり、このようなことの実現は非常に困難と考えられる。

【0009】本発明の目的は、上記の課題を解決することにあり、副側制御装置間の通信回数を減らして、ライトデータの2重化を保証する処理を効率化することである。本発明の他の目的は、ライトデータの2重化を保証する処理を1台の副側制御装置に集中させず各副側制御装置に処理を分散させることである。本発明の他の目的は、副側のマスタ制御装置を必要とせず、副システムを構成する制御装置の数が多数に及んでも実現可能（副システム構成する制御装置の数に影響されない）な、データの2重化手段を提供することである。

【0010】

【課題を解決するための手段】上記目的を達成するため、正システムを構成する1台以上の正側制御装置から副システムを構成する1台以上の副側制御装置へデータのバックアップを行う記憶装置システムにおいて、処理装置から前記正側制御装置にライト要求があった場合に、前記ライト要求のデータにライト時刻に関する時刻

情報を付与して前記正側制御装置内のキャッシュメモリに格納するステップと、前記格納された前記ライトデータと前記時刻情報を前記副側制御装置に送るステップと、前記副側制御装置内のキャッシュメモリに前記ライトデータと前記時刻情報を格納するステップと、前記副側制御装置内の前記キャッシュメモリに格納された前記時刻情報を、前記副側制御装置間を互いに接続した通信路を通して他の副側制御装置に送るステップと、第1の正側制御装置から受け取った第1の副側制御装置の時刻情報と、第2の副側制御装置から送られた時刻情報を比較して、古い方の時刻情報を前記通信路を通して次の副側制御装置に送るステップと、前記古い方の時刻情報を前記複数の副側制御装置間を一巡させて、前記複数の副側制御装置間に前記ライトデータが二重化されたことを保証するデータ保証時刻を決定するステップとを備えるようにする。他の実現方法は、発明の実施の形態の中で、詳しく説明される。

【0011】

【発明の実施の形態】図1は、本発明の実施形態におけるデータ2重化システムの構成概要図である。正側複合記憶装置システムは、複数(M台、1～m)の正側制御装置210、220、230、240から構成される。各正側制御装置は処理装置10と各々接続される。副側複合記憶装置システムは、複数(N台、1～n)の副側制御装置310、320、330、340から構成される。処理装置と正側制御装置間、正側制御装置と副側制御装置間は、それぞれインタフェースケーブル40によって接続されている。また、副側制御装置間は通信ケーブル50によってリング状に接続され、副側制御装置間で通信できる。ここで、インタフェースケーブル40と通信ケーブル50を便宜上名称を使い分けているが、どちらも情報を通信することが出来るものであれば良い。また、信頼性を高めるため、冗長化されたインタフェースケーブル40、通信ケーブル50を使用しても良い。

【0012】処理装置がライト要求を発行すると、ライト要求を受けた正側制御装置は、当該正側制御装置とインタフェースケーブルで接続された副側制御装置へ、ライト要求に含まれるデータを二重化するため、ライト要求を副側制御装置へ発行する。正側制御装置から副側制御装置へライト要求のデータがインタフェースケーブル40上で転送される場合、副側制御装置が受け取るライト要求のデータの受信順序は、ホストが正側制御装置へライト要求を発行した順序とは必ずしも一致しておらず、異なっていることがある。このため、副側制御装置では、これを正しい順序(ライト要求が発行された順序)に並べ替える必要がある。これを実現するため、正側の制御装置は、ホストからライト要求を受け付けた順にシーケンシャル番号を付与し、副側制御装置へライト要求のデータを転送する際は、このシーケンシャル番号も合せて送るようにする。副側制御装置はこのシーケン

シャル番号を参照することで、ホストがライト要求を発行した順序を知ることができる。

【0013】また、副側複合記憶装置システム内の副側制御装置間の通信手段として、複数の副側制御装置を通じケーブルでリング状に接続した場合の手段について述べているが、その他の通信手段として、副側制御装置が赤外線や無線による通信手段を用いて副側制御装置間での通信を行う手段や、副側制御装置をバス接続によって接続し、バスを用いて通信を行う手段等も考えられる。いずれの通信プロトコル(通信手段)を用いるにしろ、以下に説明する本実施例の同様の方法で(本実施例と同種の情報を通信することで)データの2重化処理を実行することが可能である。

【0014】次に、制御装置の構成を詳しく説明する。

【0015】図14は、正側制御装置210の構成図を示す。なお、本実施例における制御装置は正側、副側にかかわらず全て制御装置210と同じ構成である。制御装置は、処理装置10とのデータや制御情報の受け渡し、他制御装置111とデータや制御情報の受け渡し(正副制御装置間でのデータや制御情報の受け渡し)、及び、副側複合記憶装置システム内の複数の副側制御装置間でのデータや制御情報の受け渡しを行う、マイクロコンピュータ制御のチャネルアダプタ112を有する。更に、制御装置は、ディスクドライブ等の外部記憶装置113とのデータの受け渡しを行うマイクロコンピュータ制御のディスクアダプタ114、受領したデータを格納するための不揮発キャッシュメモリ115、制御装置の共有情報を格納するためのシェアドメモリ116、制御装置内でチャネルアダプタ、ディスクアダプタ、不揮発キャッシュメモリ、及び、シェアドメモリ間のデータや制御情報の転送を行うバス119を有する。ここで、図14の当該装置が正側制御装置の場合は他制御装置111とは副側制御装置に相当し、この2つの正副制御装置間でデータの受け渡しや正副間の制御装置間通信が行われる。また、図14の当該装置が副側制御装置の場合は、他制御装置としては正側制御装置のほか、副側複合記憶装置システム内の他の副側制御装置に相当することもある。なお、制御装置210と外部記憶装置113とを合せて、記憶システムを構成する。この記憶システムは処理装置10から見て、外部記憶装置となる。処理装置がメインフレーム系のコンピュータの場合、外部記憶装置はサブシステムと呼ばれる。

【0016】図15は、図14で示したチャネルアダプタ112の詳細な構成図である。チャネルアダプタ112は、処理装置10及び他の制御装置111からのデータを受領するデータ受領部211、処理装置10及び他の制御装置111にデータを送信するデータ送信部212、本発明を実行するためのプログラムを含むプログラムの格納部分である不揮発メモリ117、及び、プログラム実行部分であるマイクロプロセッサ118より構成される。

【0017】次に、図2および図3を用いて、ホストがライト要求を発行する際の正側制御装置の動作、および、正側制御装置から副側制御装置へライト要求を発行する際の正副制御装置の動作を説明する。

【0018】図2は、正側制御装置1(210)における、処理装置10からのライト要求を受けた場合の処理について表している。正側制御装置1(210)のデータ受領部213(図15のデータ受領部211)は、処理装置10からライト要求を受領すると、ライトデータに時刻情報を付与して、冗長化された正側不揮発キャッシュメモリ212a、212b(図14の不揮発キャッシュメモリ115)に格納し、処理装置10にライト要求の完了報告を行う。時刻情報としては、処理装置がタイマーを有し、タイマーから得た時刻情報を付与したライト要求を発行してくる場合は、このライト要求に付与されている時刻情報を用いることができる。一方、処理装置がタイマーを有していない場合は、複数の正側制御装置が有するタイマーからライト要求を受けた時刻を取得し、取得した時刻情報を用いる。なお、制御装置は、前述したように、処理装置が発行した順序、即ち、正側処理装置がライト要求を受け付けた順序でシーケンシャル番号も合せて格納する。

【0019】図3は、正側制御装置1(210)から副側制御装置1(310)へライト要求を発行した場合の処理を示す。正側制御装置1(210)のデータ送信部214(図15のデータ送信部212)は、インタフェースケーブル40で接続された副側制御装置1に対してライト要求を発行し、正側不揮発キャッシュメモリ212aに格納したライトデータ、時刻情報、シーケンシャル番号の転送を行う。副側制御装置1(310)のデータ受領部215(図15のデータ受領部211)は、受け取ったライトデータ、時刻情報(処理装置から正側へのライト時刻であり、以下に記述する「時刻情報」もすべて処理装置から正側へのライト時刻)、シーケンシャル番号を副側不揮発キャッシュメモリ312a(図14の不揮発性キャッシュメモリ115)に格納する。このほかの正側制御装置2~mと、副側制御装置2~nの間でも、処理装置からライト要求を受けた場合は、同様に、正側制御装置から副側制御装置へライト要求が発行される。

【0020】図4は、副側制御装置1(310)の副側不揮発キャッシュメモリ312a(図14の不揮発性キャッシュメモリ115)上に設けられたライトデータ管理情報テーブル700の構造を表している。副側制御装置1は、正側制御装置1(210)から受領したライトデータを副側不揮発キャッシュメモリ312aに設けられたライトデータ管理情報テーブル700によって管理する。このライトデータ管理情報テーブルには、シーケンシャル番号701、ライトデータの時刻情報721、ライトデータが格納されている副側不揮発キャッシュメモ

リ312a(図14の不揮発性キャッシュメモリ115)のアドレス情報711等が、格納される。副側制御装置310は、正側制御装置210から送られるライトデータを、必ずしも時刻情報の古いものから順に受領するとは限らず、ライトデータに一時的に抜けが発生する場合がある。そのため、副側制御装置は、ライトデータ管理情報テーブルに格納されているシーケンシャル番号701を参照し、この番号に抜けが有るか否かを調べ、このシーケンス番号に抜けが無く連続して受領している最新のライト要求を見つける。そして、副側制御装置は、このライト要求が有する時刻情報が示す時刻までのライトデータを受信していると判断する。即ち、副側制御装置310は、ライトデータ管理情報テーブル700を参照することにより、どの時刻以前の時刻情報を持つライトデータは連続して受領済みで、どの時刻よりも後の時刻情報を持つライトデータには抜けがあるかを確認することが出来る。テーブル700の例では、シーケンシャル番号705に抜けがあり、シーケンシャル番号がNo.5であるライトデータを未受領であり、時刻情報4以前の全ライトデータを、副側制御装置が受領していることが分かる。

【0021】副側複合記憶装置システムの各々の副側制御装置は、このライトデータ管理情報テーブルを有し、自身が受領したライトデータ及びその時刻情報の管理を行っている。また、1台の副側制御装置に対して、複数台の正側制御装置が接続されるケースでは、副側制御装置は接続される正側制御装置の台数分のライトデータ管理情報テーブルを持つ。

【0022】次に、正側制御装置からライトデータを受領した副側制御装置が保証時刻を決定する動作と保証時刻を通知する動作の説明を通して、受領したライトデータの2重化を保証する方法を以下に説明する。最初に図16を用いて、保証時刻の決定と保証時刻の通知のプロセスを行うために副側制御装置間で巡回される情報について説明しておく。図16は、副側制御装置間で巡回される情報を示しており、制御装置内では副側不揮発キャッシュメモリ312a上に格納される。各情報の詳細・具体的な使われ方などは後述する。Serial#2001は、当該制御装置を識別するための制御装置の固有情報であり、副側不揮発キャッシュメモリ312a上に格納されるほか、各制御装置のシェアドメモリ116にも格納されている。要求種別2002には、「要求種別1=保証時刻決定」2003及び「要求種別2=保証時刻通知」2004の2つがあり、要求種別2002を受領した副側制御装置が「保証時刻決定の処理を行うのか」、「保証時刻通知の処理を行うのか」を判断するために使用される情報である。保証時刻2005は、それ以前の時刻情報を持つライトデータについては全て副側制御装置で受領済みであることを保証する時刻である。仮保証時刻2006は、保証時刻決定のプロセスにおいて、正式な保証時刻2005が決定されるまでの仮の保

証時刻であり(詳細は後述)、これらの情報は副側不揮発キャッシュメモリ312a上に格納される。

【0023】(1) 保証時刻を決定する処理

正側制御装置から受け取ったライトデータの2重化を保証する保証時刻の決定する際の各副側制御装置での処理プロセスを以下説明する。なお、この処理プロセスは、図6に示す副側複合記憶装置システムを例にして説明する。ここで、例とする図6のシステムは、データ受領部(313、323、333、343)とデータ送信部(314、324、334、344)とを有するn個の副側制御装置(310、320、330、340)が、通信ケーブル50によってリング状に接続される構成を有するものである。なお、データ受領部(313、323、333、343)は図15のデータ受領部211が対応し、データ送信部(314、324、334、344)は図15のデータ送信部212が対応する。

【0024】また、基準時刻を説明しておく。副側制御装置1(310)は、ライトデータの2重化を保証するために、副側制御装置1のライトデータ管理情報テーブル700を参照し、どの時刻までのライトデータが抜けがなく、正側制御装置から連続して到着しているかをシーケンシャル番号を用いて確認し、連続して受領しているライトデータに付与されている時刻を「基準時刻」として副側不揮発キャッシュメモリ312aに記憶する。このようにして決定された基準時刻は、当該副側制御装置に接続された正側制御装置から送信された、それ以前の時刻情報を持つライトデータについては全て当該副側制御装置で受領済みであると保証できる時刻であると言える。ライトデータ管理情報テーブル700は、各副側制御装置が有しており、基準時刻は各副側制御装置ごとに決定され、各副側制御装置の副側不揮発キャッシュメモリに記憶される。

【0025】以下、具体的に保証時刻決定のプロセスを説明する。

【0026】まず、最初に保証時刻を決定するための情報を送信する副側制御装置の動作を説明する。副側制御装置1(310)は、副側制御装置間通信送信部314により、シェアドメモリ116上に格納されている自身のSerial#(各制御装置に固有の情報)、要求種別=保証時刻決定(送信の要求種別を保証時刻決定と云う種別にする)、及び、仮保証時刻(自身が記憶した基準時刻)2006を、接続先の副側制御装置2(320)に転送する(図6のデータ送信部314からデータ受領部323へ向けての送信)。要求種別には送り側の副側制御装置1(310)が何を目的にして送信を行っているかの情報が格納されており、受け側の副側制御装置2(320)は要求種別を参照してどういう処理をすれば良いのかを判断することができる。

【0027】次に、図5を用いて、副側制御装置が送信した情報を順次受信する副側制御装置の動作を説明する。副

側制御装置2(320)は、副側制御装置1から、要求種別1=保証時刻決定をデータ受領部(323)で受け取ると(ステップ802)、受信したSerial#とシェアドメモリに格納されている自Serial#とを比較して、受信したSerial#2001と自身のSerial#が異なるか否かを確認する(ステップ803)。異なる場合、副側制御装置2(320)は、自身の基準時刻を決定し(ステップ804)、前段の副側制御装置1(310)から受け取った仮保証時刻2006と比較する(ステップ805)。副側制御装置2(320)は、比較した結果、自身の基準時刻の方が受領した仮保証時刻2006よりも過去ならば、仮保証時刻2006を自身の基準時刻で更新する(ステップ806)。そして、副側制御装置2は、ステップ801において、副側制御装置1(310)から受け取った副側制御装置1(310)のSerial#2001、要求種別1=保証時刻決定2003とともにステップ806で更新した仮保証時刻を、次副側制御装置3(330)に送信する(ステップ807)。逆に、ステップ805で比較した結果、受領した仮保証時刻2006の方が古いならば受領した仮保証時刻2006は更新せず、副側制御装置1(310)のSerial#2001、要求種別1=保証時刻決定2003とともにステップ802で受領した仮保証時刻をそのまま次副側制御装置3(330)に送信する(ステップ807)。

【0028】上記の処理を全副側制御装置間で繰り返し、ステップ807において最後に副側制御装置n(340)のデータ送信部(343)から副側制御装置1(310)のデータ受領部(314)へ、仮保証時刻2006、副側制御装置1(310)のSerial#2001、要求種別1=保証時刻決定2003を送信する。ステップ802において副側制御装置1(310)が、副側制御装置n(340)より受領したSerial#2001を、ステップ803において自身のSerial#と同一であることを確認することによって、受領した仮保証時刻2006が全副側制御装置を一巡して決定されたものであると認識し、仮保証時刻2006を保証時刻2005として決定する。そして、副側制御装置1は、それ自身の不揮発キャッシュメモリ(312a)に、この保証時刻を格納する。

【0029】このようにして決定された保証時刻2005はそれ以前の時刻情報を持つ、全正側制御装置から送信されたライトデータを全て副側制御装置で受領済みであると保証できる時刻である。

【0030】また、1台の副側制御装置に複数台の正側制御装置が接続されるケースでは、各副側制御装置の基準時刻は以下のように決定される。その他の保証時刻決定のプロセスは上記と全く同様に行われる。副側制御装置は接続される各正側制御装置に対応する台数分のライトデータ管理情報テーブル700を持ち、各正側制御装置に対応するライトデータ管理情報テーブル700を参照することで、各正側制御装置からどの時刻までのライトデータが抜けがなく、連続して到着しているかを確認することが出来る。その時刻を仮基準時刻と定義すると、副側制御装置は各正側制御装置に対応する仮基準時刻の中

から1番古い時刻を持つ仮基準時刻を基準時刻として副側不揮発キャッシュメモリに格納する。このようにして決定された基準時刻は、当該副側制御装置に接続された複数の正側制御装置から送信された、それ以前の時刻情報を持つライトデータは全て当該副側制御装置で受領済みであると保証できる時刻であると言える。

【0031】(2) 保証時刻を通知する処理

上記で保証時刻決定の処理について述べ、保証時刻2005が、保証時刻以前の時刻情報を持つ全ライトデータの2重化が保証されていることを示すことを説明したが、各副側制御装置は決定された保証時刻2005を知らないため、どの時刻までの時刻情報を持つライトデータの2重化が保証されているのかを判断することが出来ない。決定された保証時刻2005を全副側制御装置に通知し、各副側制御装置がどの時刻までのライトデータの2重化が保証出来るのか知ることで、各副側制御装置は、保証時刻2005以前の時刻情報を持つライトデータの2重化を保証できる。

【0032】まず、保証時刻を決定し不揮発キャッシュメモリにその保証時刻を格納した副側制御装置1(310)は、副側制御装置間で巡回される情報のうち、要求種別2002を「要求種別=保証時刻通知」2004(送信の要求を「保証時刻通知」という種別にする)に更新し、決定した保証時刻2005、自身のSerial#2001とともに更新した要求種別(保証時刻通知)を副側制御装置2(320)に送信する。

【0033】次に、図7を用いて、保証時刻通知を要求種別にもつ巡回情報を受領した副側制御装置の動作を説明する。ステップ902において、「要求種別=保証時刻通知」をデータ受領部(323)で受け取った副側制御装置2(320)は、受け取ったSerial#2001と自身のSerial#が異なるか否かを判定する(ステップ903)。異なる場合、受領した保証時刻2005を自身の不揮発キャッシュメモリに格納する(ステップ904)とともに、ステップ902で受信した保証時刻2005を、要求種別2=保証時刻通知2004、副側制御装置1(310)のSerial#2001とともに副側制御装置3(330)に送信する(ステップ905)。

【0034】上記の処理をリング状に接続された通信ケーブル50に沿って全副側制御装置間で繰り返し、最後にステップ901において副側制御装置n(340)のデータ送信部(344)から副側制御装置1(310)のデータ受領部313に保証時刻2005、副側制御装置310のSerial#2001、要求種別=保証時刻通知2004を送信する(ステップ905)。そして、副側制御装置1(310)は、ステップ902において副側制御装置n(340)からSerial#を受領し、ステップ903において、受領したSerial#2001が自身のSerial#と同一であることを確認すると、保証時刻通知がリング状に接続された副側制御装置を一巡したとして、保証時刻2005が全副側制御装置に通知された

ことを認識する。

【0035】以上(1)(2)で説明したように、保証時刻2005を決定し全副側制御装置に通知することによって各副側制御装置は自身の副側不揮発キャッシュメモリ上のライトデータが保証時刻2005以前のものか以後のものかを判断することが可能となり、保証時刻2005以前の時刻情報を持つライトデータの2重化が保証される。

【0036】(3) 保証時刻の決定と保証時刻の通知と一緒にを行う処理

次に、保証時刻の通知と、次の保証時刻の決定とと一緒にして同時的に行う場合の各副側制御装置での処理プロセスを示す。次の保証時刻の決定と保証時刻の通知と一緒にを行うことにより、処理の効率化を図ることができる。また、従来の方法では、データの2重化を保証するために、保証時刻の決定のための通信と保証時刻の通知のための通信とでマスタ制御装置と各副側制御装置の間で2回の通信が必要であったのに対し、本処理方法では保証時刻の決定と保証時刻の通知のための通信と一緒に(同時的に)行うことにより、各副側制御装置間で1回の通信を行えばデータの2重化を保証することができる。

【0037】まず、副側制御装置1(310)は、副側制御装置n(340)から受け取った仮保証時刻2006を保証時刻2005として決定し、自身の不揮発キャッシュメモリ312aに格納するとともに、保証時刻2005を要求種別2=保証時刻通知2004とともに副側制御装置2(320)に転送する。この際、更に、副側制御装置1(310)は、次の保証時刻決定のための自身の現在の基準時刻である仮保証時刻2006、自身のSerial#2001、要求種別1=保証時刻決定2003も合せて、副側制御装置間通信送信部314より、接続先の副側制御装置2(320)に送信する。即ち、自制御装置のSerial#、要求種別1(保証時刻決定)、仮保証時刻、要求種別2(保証時刻通知)、保証時刻が、次副側制御装置へ、一緒に転送される。

【0038】次に、保証時刻通知と保証時刻決定要求とを1つの処理として受領した副側制御装置の動作を、図8で説明する。副側制御装置2(320)は、要求種別1=保証時刻決定2003と要求種別2=保証時刻通知2004をデータ受領部(323)で一緒に受け取ると(ステップ1002)、受け取ったSerial#と自身のSerial#が異なることを確認し(ステップ1003)、受領した保証時刻2005を自身の不揮発キャッシュメモリに格納する(ステップ1004)。更に、副側制御装置2は、自身の基準時刻を決定し、副側制御装置1(310)から受け取った仮保証時刻2006と比較する(ステップ1005)。比較した結果(ステップ1006)、副側制御装置2自身の基準時刻の方が仮保証時刻2006よりも過去ならば、ステップ1007において仮保証時刻2006を自身の基準時刻で更新し、ステップ1001へ進む。過

去でないなら、受信した仮保証時刻をそのままにし、その仮保証時刻は更新せずに、ステップ1008へ進む。そして次のステップ1008において、副側制御装置1(310)から受け取った副側制御装置1(310)のSerial#2001、要求種別1=保証時刻決定2003を仮保証時刻とともに副側制御装置3(330)に送信するのと一緒に、要求種別2=保証時刻通知2004と保証時刻2005も次段の副側制御装置3(330)に送信する。

【0039】上記の処理を全副側制御装置間で繰り返し、ステップ1002において最後に副側制御装置n(340)のデータ送信部(343)から、副側制御装置1(310)のデータ受領部(314)へ、副側制御装置1(310)のSerial#2001、要求種別1=保証時刻決定2003、仮保証時刻2006、要求種別2=保証時刻通知2004、保証時刻2005が送信される。副側制御装置1(310)は、副側制御装置n(340)からSerial#2001を受信し(ステップ1002)、この受信したSerial#が自身のSerial#2001と同一であることを確認することによって(ステップ1003)、保証時刻2005が全副側制御装置に通知されたことを認識する。更に、受領した仮保証時刻2006が全副側制御装置を一巡して決定されたものであると認識し、仮保証時刻2006を次回の保証時刻として決定する。

【0040】以上のように保証時刻の決定処理と保証時刻の通知処理とを一緒に行うことにより、ライトデータの2重化を保証するために必要な、各副側制御装置間での通信回数を1回に抑えることが可能となる。

【0041】次に、リング状に接続された副側制御装置間で、保証時刻の決定と保証時刻の通知を一緒に行うことによる処理を複数個(例では2個)並行して実行した場合について説明する。処理を複数個並行して実行することにより、ライトデータの2重化を保証する処理の実行間隔を短縮することが可能となる。図9および図10は、これを説明する図である。図9は、リング状に接続された副側制御装置間で、上記の保証時刻決定と保証時刻の通知を一緒に行うことの概念図を示す。図10は、図9に示す方法の優位性を示すための比較例を示すものであり、上記の保証時刻決定と保証時刻の通知処理を1つだけ行った場合の概念図を示す。

【0042】図9において、副側制御装置間で巡回する情報1(1101)は、処理(3)で説明した保証時刻決定と保証時刻通知を一緒に行うための情報を含んでおり、Serial#2001、要求種別1=保証時刻決定2003、要求種別2=保証時刻通知2004、保証時刻2005、仮保証時刻2006から構成される。また、情報2(1102)も副側制御装置間で巡回する、保証時刻決定と保証時刻通知のための情報(情報1と同種の情報だが中身が異なる)である。そして、図9では、これら情報1(1101)と情報2(1102)の2つを、リング状に接続された副側制御装置間(310, 320, 330, 340)で並行して巡回させることによって、保証時刻の

決定と保証時刻の通知処理が2つ並行して行われるようになる。これは、2つの情報が同時に巡回しているので、1つの場合と比べて、情報が巡回してくる間隔が短く、保証時刻の決定と保証時刻の通知が速く行われるようになる。

【0043】図10では、1つの情報1だけがリング状に接続された副側制御装置間で巡回しているので、図9のように情報を2つ同時に巡回させている場合と比較して、情報が巡回してくる間隔が長くなり保証時刻の決定と保証時刻の通知が遅くなる。

【0044】今、図9および図10において、リング状に接続された4つの副側制御装置間を情報が一巡して保証時刻以前の時刻情報を持つライトデータの2重化が保証されるまでの時間を間隔T(1201)とする。

【0045】この場合、図10に示す例では、ライトデータの2重化の保証処理を1つの情報1の巡回を繰り返して保証するので、図11に示すように間隔Tでライトデータの2重化保証処理が繰り返されることになる(1201, 1202, ...)。見方を変えれば、現時刻よりも間隔T以前のライトデータは、2重化が保証されていることになる。

【0046】一方、図9に示す例では、リング状に接続された複数の副側制御装置間で、情報1(1101)と情報2(1102)を、T/2の間隔で2個並行して巡回することになる。そうすると、情報1が副側制御装置間を一巡し保証時刻以前の時刻情報を持つライトデータの2重化を保証した後、T/2の間隔で、情報2が副側制御装置間を一巡し、保証時刻以前の時刻情報を持つライトデータの2重化を保証することになる。即ち、ライトデータの2重化保証処理を2つの情報(情報1、情報2)の巡回を繰り返した場合、図12に示すように間隔T/2でライトデータの2重化保証処理が繰り返される(1301, 1302, 1303, 1304, ...)。見方を変えれば、現時刻よりも間隔T/2以前のライトデータは2重化が保証されていることになる。

【0047】このように、リング状に接続された副側制御装置間で、保証時刻の決定と保証時刻の通知を一緒に行うことの処理を1個だけ実行した場合(図10の場合)に比べ、リング状に接続された副側制御装置間で、保証時刻の決定と保証時刻の通知を一緒に行うことの処理を複数個(図9は2つの場合)並行して実行することによって、より最近のライトデータまで2重化を保証することができる。

【0048】上述の説明では、リング状に接続された副側制御装置間で、保証時刻の決定と保証時刻の通知を一緒に行うことの処理を2個並行して実行した場合の例にとったが、処理を3個、4個と複数個並行して実行することにより、さらに最近のライトデータまで2重化を保証することも可能となる。但し、ここで注意しておきたいのは、保証時刻の決定と保証時刻の通知を一緒に行うことの情

報が、副制御装置間をリング状に結ぶ通信ケーブル上を、物理的に同時に複数通信が必要である。もし、副制御装置間が論理的にリング状に結ばれている場合でも、保証時刻の決定と保証時刻の通知を一緒に行う処理のための情報が通信ケーブルの上で、同時に転送できなければ、保証間隔Tを短縮することは難しい。例えば、バスを用いて論理的に巡回接続を実施した場合には、負荷が集中する副制御装置は存在しないが、バスは同時に1つの通信しかできないので、保証時刻の決定と保証時刻の通知を一緒に行う処理を2個並行して送信することはできない。

【0049】また、2重化の保証された副側不揮発キャッシュメモリのライトデータは、外部記憶装置113に書き込まれた後、副側不揮発キャッシュメモリから破棄されることが許される。そのため、ライトデータの2重化が速く保証されれば、外部記憶装置にライトデータを速く書き込み、不揮発性キャッシュメモリからライトデータを速く破棄できるので、キャッシュメモリの使用率を抑えることができ、システムとしての動作性能を高めることができる。即ち、保証時刻決定及び保証時刻の通知処理を複数個並行して行うことにより、データの2重化を保証するまでの時間を短縮できるので、データ2重化処理に伴う副側不揮発キャッシュメモリの使用率を低くすることが可能であるとともに、万が一、副システムに障害が発生した場合にロストするデータ量（2重化が保証できないデータ量）も少なくすることが可能となる。

【0050】図13はリング状に接続された副側制御装置（310、320、330）によって構成される副システムに、オンラインで副側制御装置4（340）を1台増設する場合の説明図である。まず、既存の通信路a（1401）（実線で示される部分）とは別に、予め、通信ケーブル50を使って、副側制御装置4（340）を1台増設した後の通信路b（1402）（破線で示される部分）を構築しておく。そして、副制御装置3（330）のデータ送信部に、通信ケーブル50で接続される2台の副側制御装置1（310）および副側制御装置4（340）の中から選択的に1台の副側制御装置を指定して情報を送信する機能を設け、指示（コマンド）により、副側制御装置3の情報送信先を副側制御装置1（310）から副側制御装置4（340）に切り替えるようにする。これにより、オンラインでの副側制御装置の増設が可能となる。

【0051】以上説明したように、本発明の実施形態に係る記憶装置システムは、次のような構成上の特徴と機能を有するものである。

【0052】本実施形態では、全ての副側制御装置（N台）を通信ケーブル等の通信路で接続し、各副側制御装置が接続先の制御装置と通信することによって、全副側制御装置間で情報が巡回可能な構成になっている。このため、各副側制御装置は、自身が正側制御装置から受け取

ったライトデータに付与された時刻情報を上記の通信路に沿って、全副側制御装置間で巡回することで保証時刻を決定し、決定した保証時刻に関する情報を上記の通信路に沿って全副側制御装置に巡回して通知することによってデータの2重化を保証することができる。なお、情報を巡回させるための通信路に関し、実施例では、リング状に接続した通信ケーブルとしたが、通信路として赤外線や無線などの通信手段を用いた場合でも、情報を副側制御装置間で巡回可能とする通信手段であればよい。また、副側制御装置をバスに接続し、バスを利用した副側制御装置間の通信により情報を巡回できるような通信手段でもよい。

【0053】また、この記憶装置システムでは、データの2重化を保証するための処理・通信は、個々の副側制御装置が分担して行い、1台の副側制御装置に処理・通信が集中するのを防いでいる。

【0054】また、このシステムでは、副側制御装置の数が何台であろうと、個々の副側制御装置には、情報の通信元及び通信先の2台の制御装置と通信できる機能が備わっていれば良い。

【0055】また、各副側制御装置が時刻情報（次の保証時刻決定のための時刻情報）と、決定した保証時刻に関する情報を同時に送・受信することにより、データの2重化を保証する処理における、各副側制御装置間での通信回数を1回に抑えることができる。

【0056】また、通信路で接続された2台の副側制御装置a、bの間に副側制御装置cを1台増設する場合に、増設する副側制御装置cを2台の副側制御装置a、bと接続して予め通信路を構築し、副側制御装置aの通信先を副側制御装置bから副側制御装置cに切り替えること（1403）により、オンラインでの制御装置の増設を可能としている。

【0057】また、M対N台の複数の正・副制御装置間で、データの2重書きを行う機能に関して、従来の方法ではデータの2重化を保証するためにはマスタ制御装置とその他の各副側制御装置の間で保証時刻の決定と通知とで各々計2回の通信回数が必要であるが、本実施例では、各副側制御装置間で1回の通信で保証時刻の決定と通知とを同時に行うことと可能とし、データ2重化処理の効率化を図り、正・副の制御装置の数が多数に及ぶ場合においても、1台の副側制御装置に負荷を集中させることなく、データの2重化を行えることを示している。

【0058】

【発明の効果】全副側制御装置をループ接続し、各副側制御装置は、正側制御装置から受信したデータに付与された時刻を前記ループ通信路で巡回することで保証時刻を決定できる。また、決定した保証時刻をループ通信路で巡回することで、保証時刻を通知できる。また、保証時刻の決定および通知を通して、複数の正側制御装置から、複数の副側制御装置へ、一貫性のとれたデータの非

同期コピーを可能とする。

【図面の簡単な説明】

【図1】本発明の実施例に係るデータ2重化システムの概略を示す。

【図2】処理装置からのライト要求に対する正側制御装置の流れを示す。

【図3】正側制御装置からのライト要求に対する副側制御装置の流れを示す。

【図4】副側制御装置に設けられるライトデータ管理情報テーブルの構成例を示す。

【図5】副側制御装置での保証時刻決定処理のフローチャートを示す。

【図6】本実施例に係る、通信ケーブルで接続された複数の副側制御装置の接続関係を示す。

【図7】副側制御装置における保証時刻の通知処理のフローチャートを示す。

【図8】副側制御装置において、保証時刻の決定処理と保証時刻の通知処理を一緒にした処理のフローチャートを示す。

【図9】保証時刻の決定処理及び保証時刻の通知処理を一緒にした処理を、2つ並行して行った場合の説明図である。

【図10】保証時刻の決定処理と保証時刻の通知処理を一緒にした処理を、1つ行った場合の説明図である。

【図11】保証時刻の決定処理及び保証時刻の通知処理を一緒にした処理を1つ行った場合の、処理の実行間隔を示す。

【図12】保証時刻の決定処理及び保証時刻の通知処理

と一緒にした処理を、2つ並行して行った場合の、処理の実行間隔を示す。

【図13】副システムに副制御装置を増設する場合の説明図である。

【図14】制御装置の構成図を示す。

【図15】チャネルアダプタの構成図を示す。

【図16】保証時刻を決定処理、及び保証時刻の通知処理に使用する情報を示す。

【符号の説明】

10 処理装置

210, 220, 230, 240 正側制御装置

310, 320, 330, 340 副側制御装置

40 インタフェースケーブル

50 通信ケーブル

212a, 212b 正側不揮発キャッシュメモリ

213 正側ライトデータ送信部

215 副側ライトデータ受領部

312a 副側不揮発キャッシュメモリ

700 ライトデータ管理情報テーブル

313, 323, 333, 343 副側制御装置間通信受領部

314, 324, 334, 344 副側制御装置間通信送信部

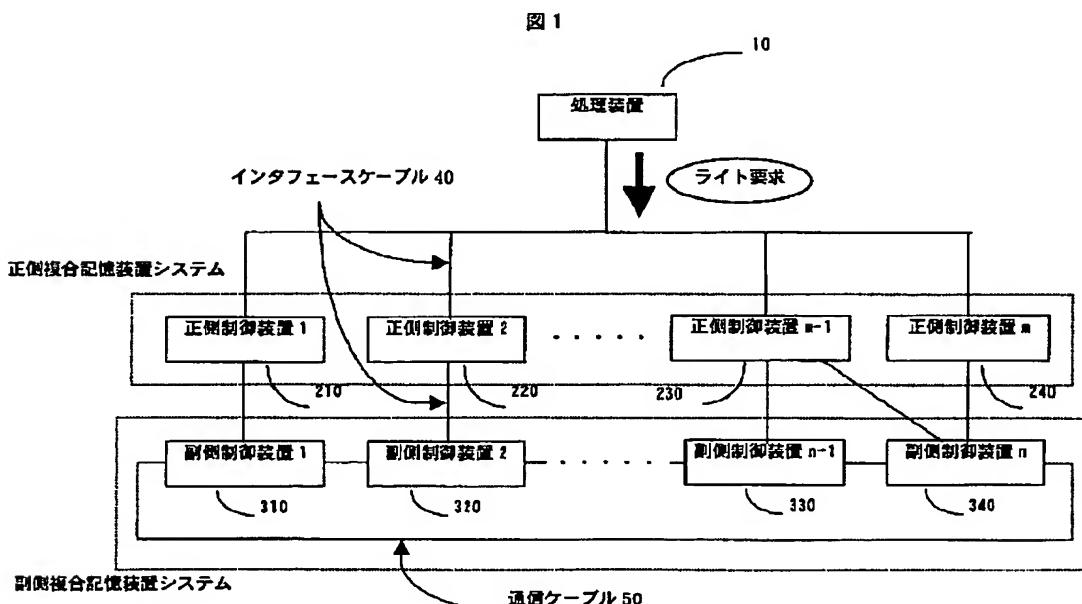
1201, 1202 2重化保証処理実行間隔T

1301, 1302, 1303, 1304 2重化保証処理実行間隔T/2

1401 通信路a

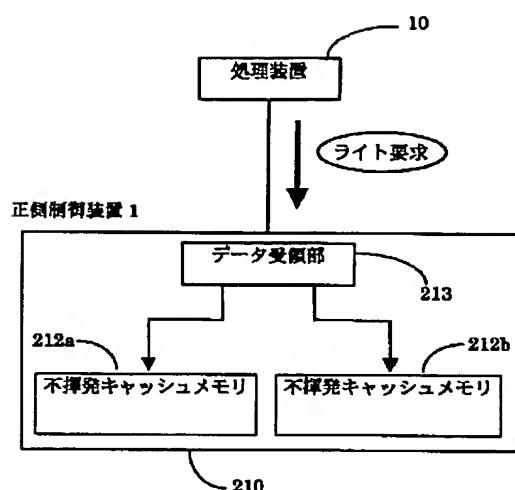
1402 通信路b

【図1】



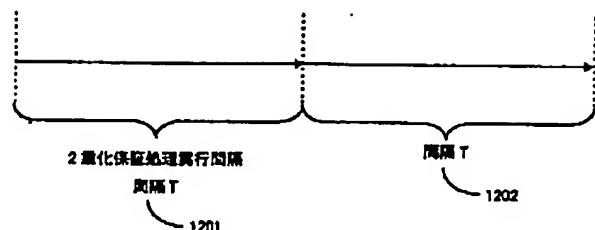
【図2】

図2



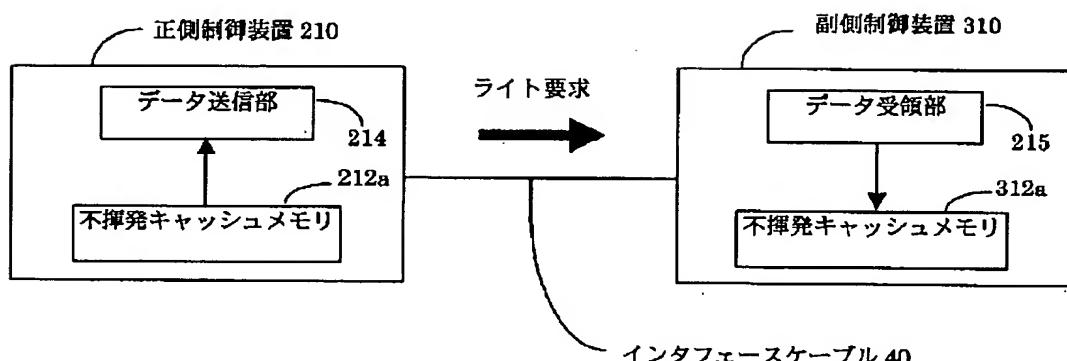
【図11】

図11



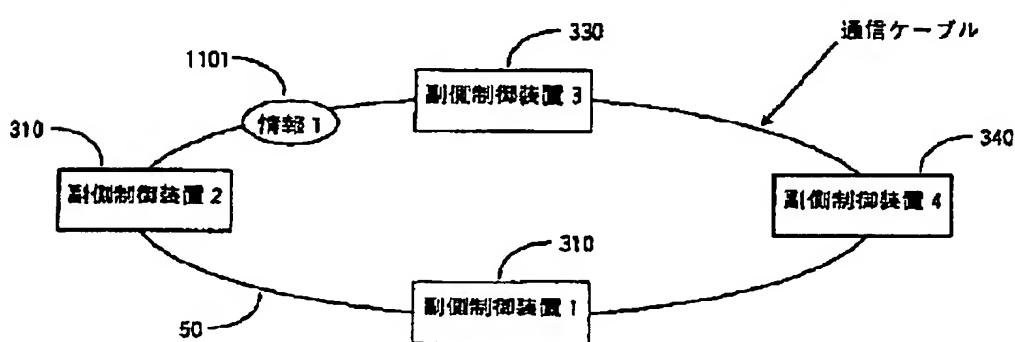
【図3】

図3



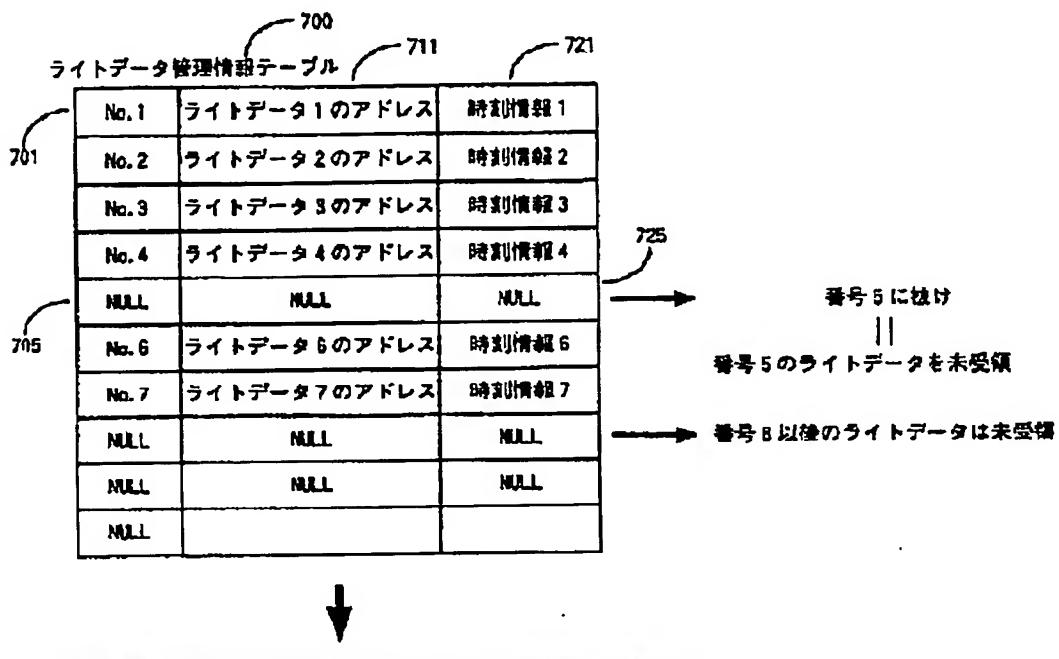
【図10】

図10



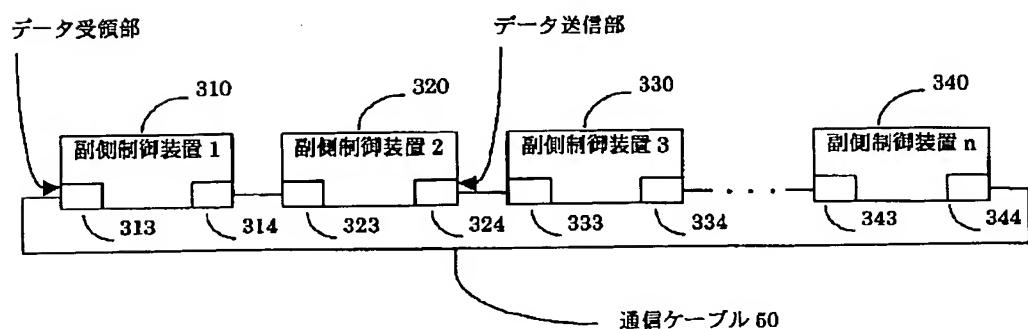
【図4】

図4



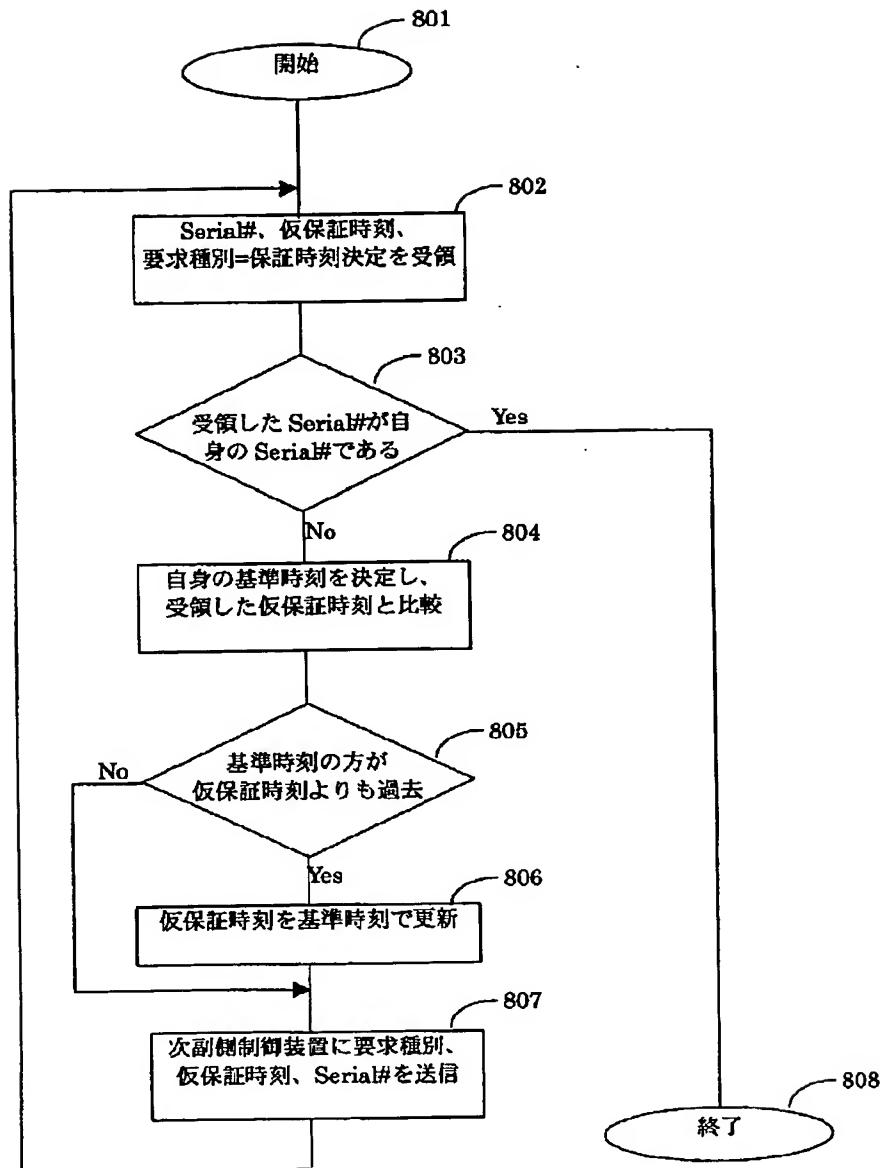
【図6】

図6



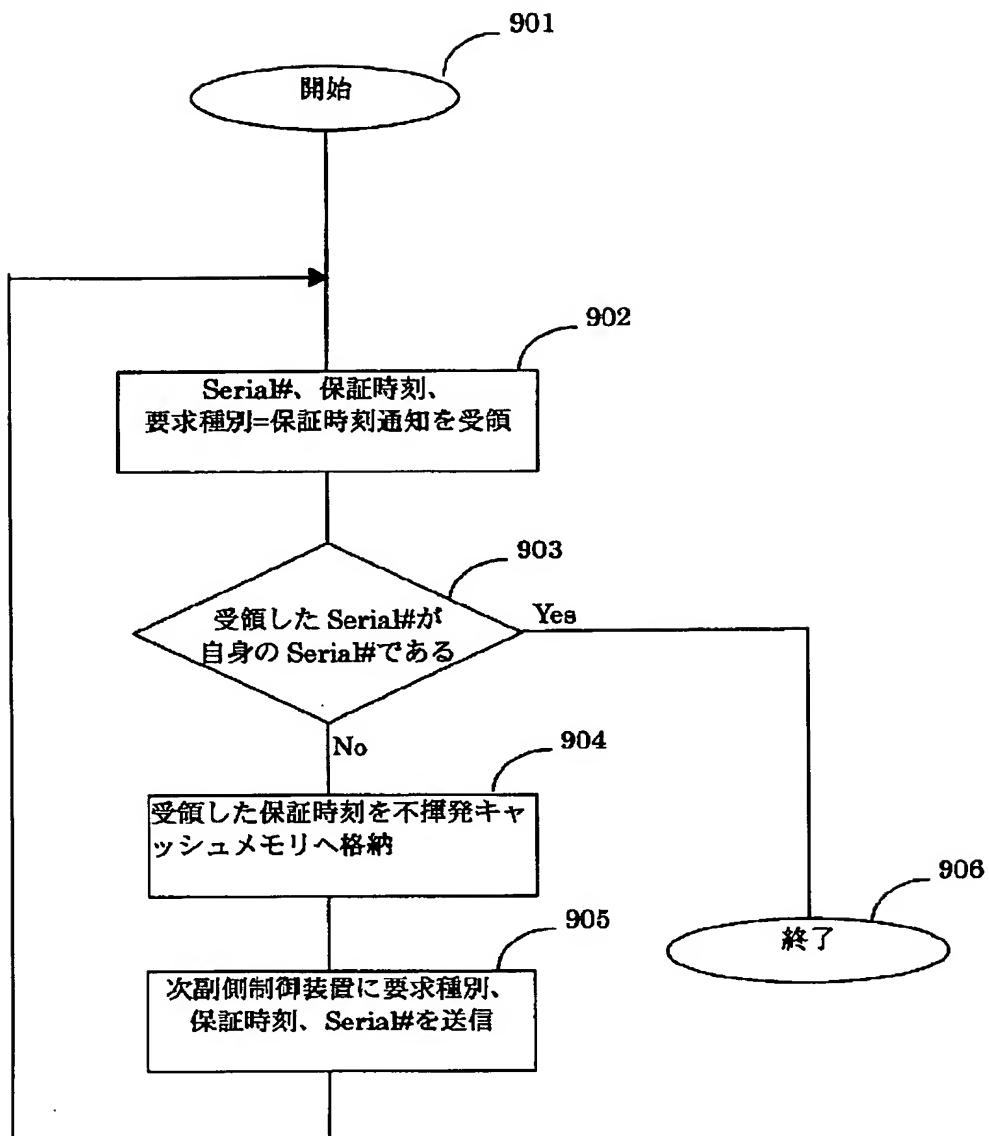
【図5】

図5



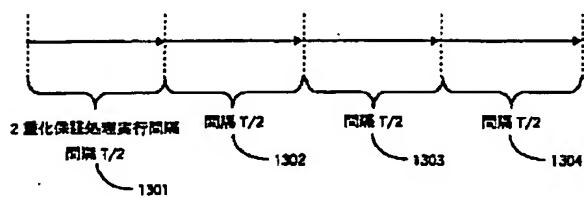
【図7】

図7



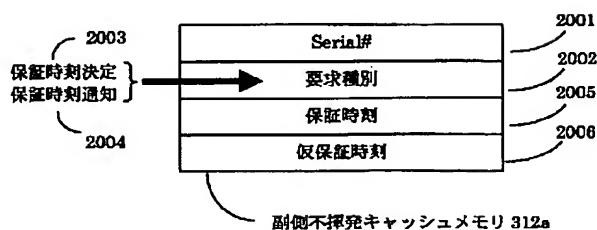
【図12】

図12



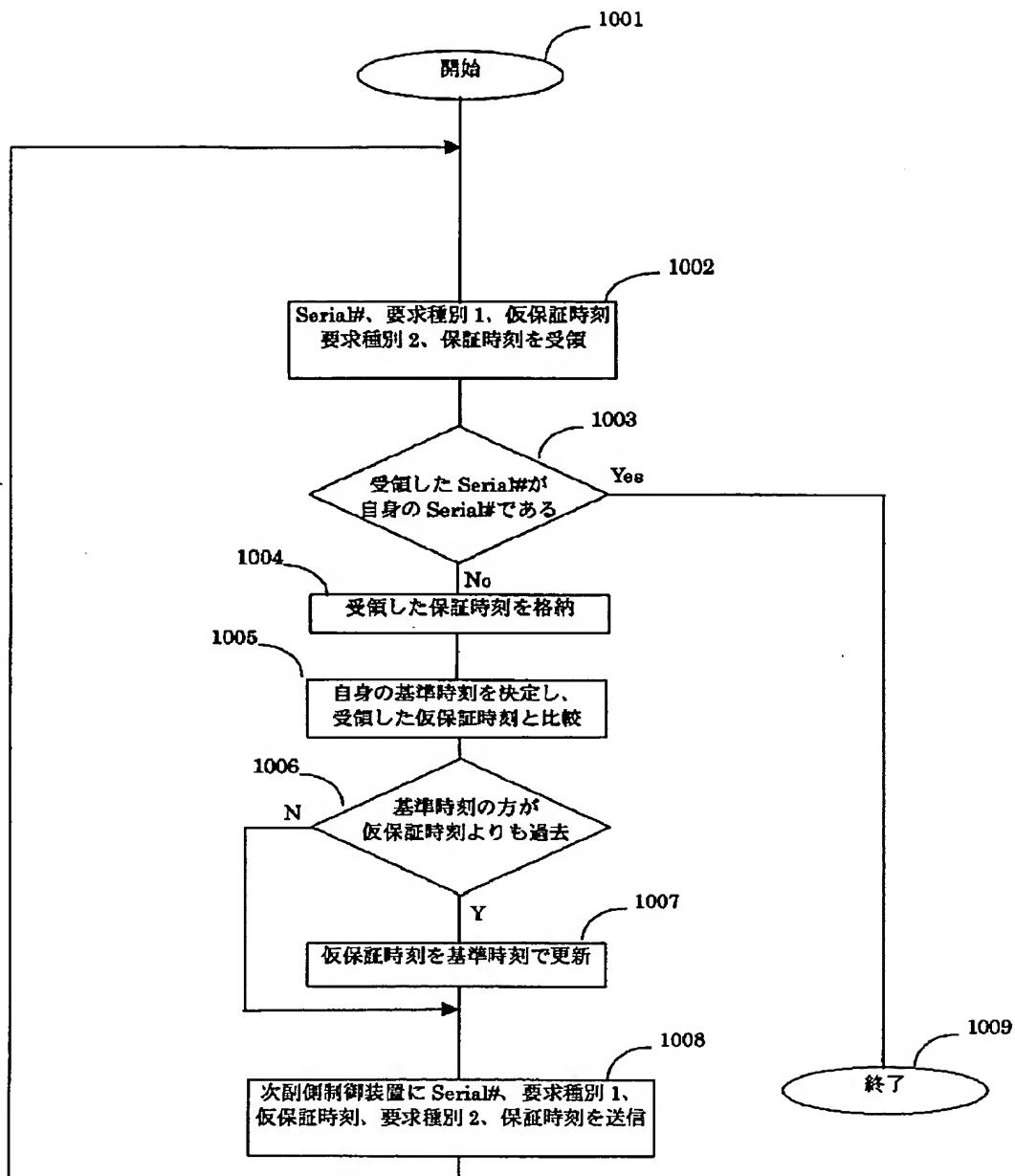
【図16】

図16



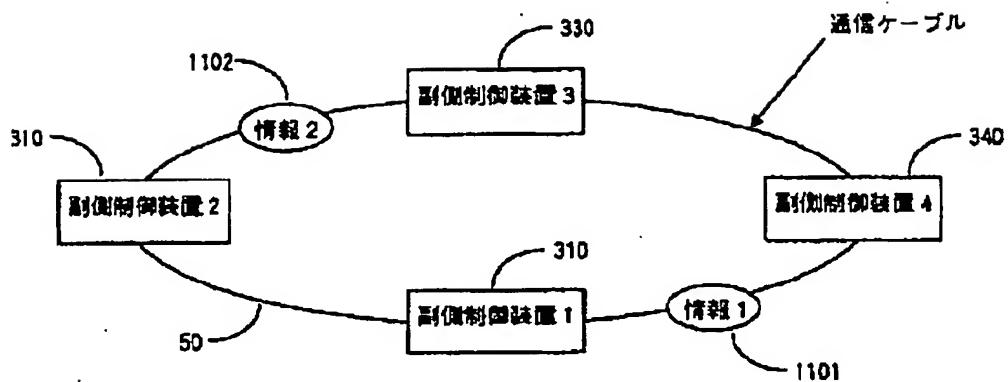
【図8】

図8



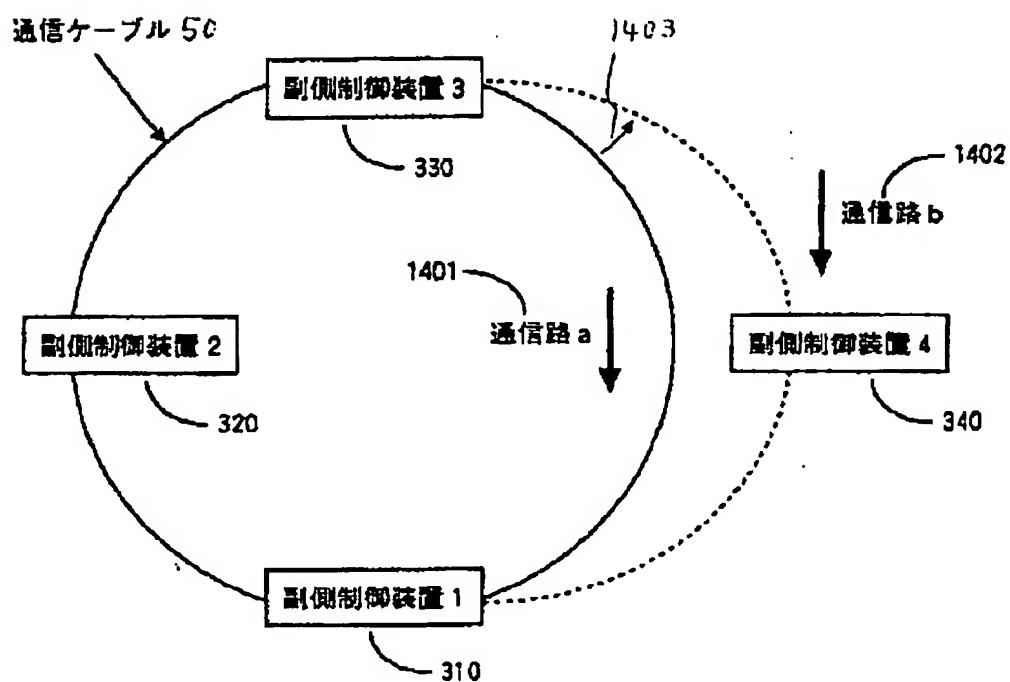
【図9】

図9



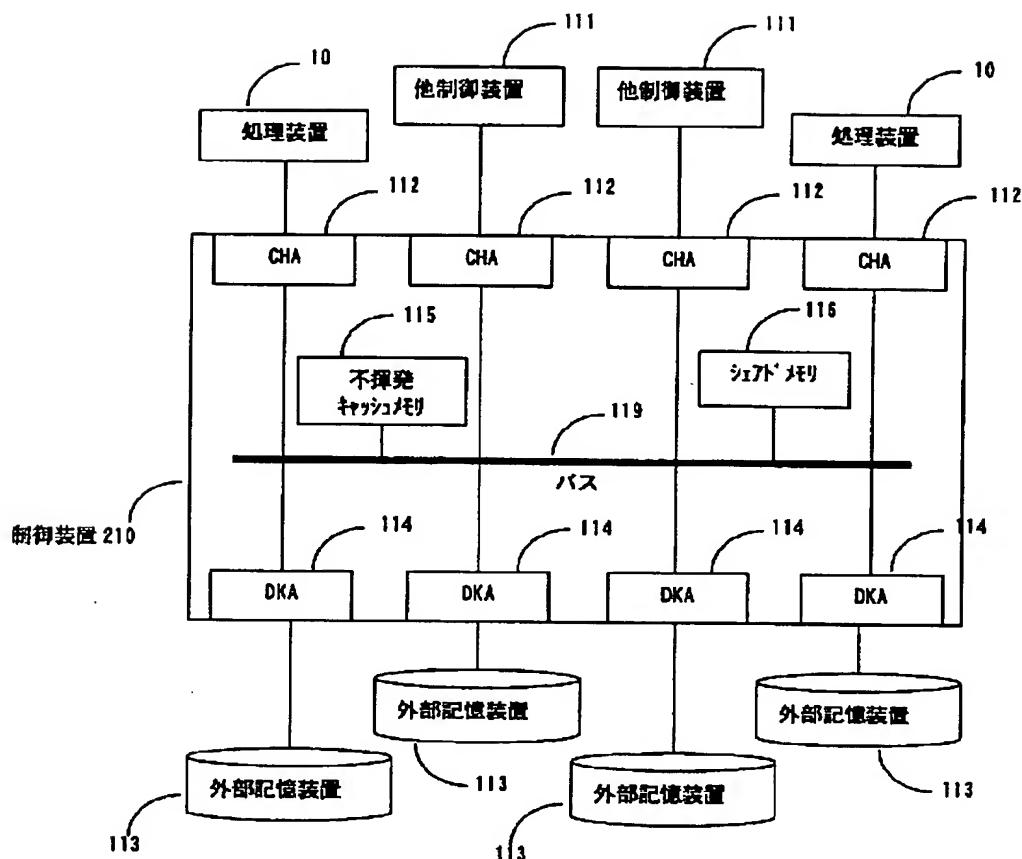
【図13】

図13



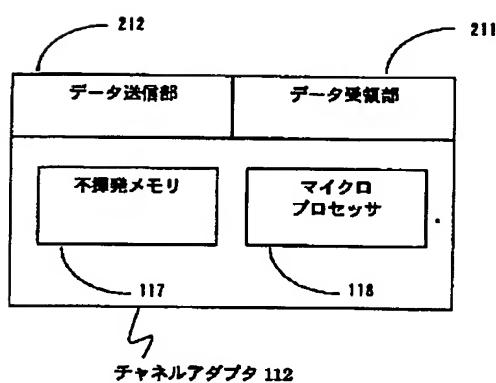
【図14】

図14



【図15】

図15



フロントページの続き

(72)発明者 長澤 光男
神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内
(72)発明者 木村 恭久
神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内

(72)発明者 小出 雄
神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内
Fターム(参考) 5B018 GA04 HA04 MA03 NA06 QA15
5B065 BA01 CA12 CE11 CH01 EA35